

Efficient semi-supervised estimation of optimal individualized treatment regimes with survival outcome

Statistical Methods in Medical Research
2026, Vol. 35(8) 1620–1636
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09622802261460255
journals.sagepub.com/home/smm



Xintong Li^{1,2,#} , Minghao Chen^{3,#}, Jialiang Li⁴ and Yong Zhou²

Abstract

Individualized treatment regimes (ITRs) represent decision-making frameworks that tailor treatment assignments to individual patient characteristics. The value function of an ITR quantifies the expected outcome under a counterfactual scenario in which such a treatment rule is applied. However, estimating optimal ITRs for survival data remains a significant challenge when outcomes are right-censored and only a subset of patients has complete outcome information due to time and cost constraints. To overcome this challenge, we formulate the problem within a semi-supervised learning framework and adopt an induced missingness perspective to model partially observed survival outcomes. We propose an imputation-based semi-supervised approach that is robust and adaptable to various imputation models. Specifically, we employ a flexible single-index kernel smoothing imputation technique to effectively utilize unlabeled data in multidimensional covariate settings. The proposed estimators for the parameters indexing the optimal ITRs are shown to be consistent and asymptotically normal. Moreover, semi-supervised estimation enhances efficiency by reducing asymptotic variance relative to supervised estimation. Numerical experiments on both simulated and real datasets demonstrate the superior performance of our proposed semi-supervised approach.

Keywords

optimal individualized treatment regimes, imputation, survival analysis, semi-supervised inference, Kernel smoothing

Received: 8 July 2025; accepted: 2 June 2026

1 Introduction

Precision medicine addresses a fundamental challenge in healthcare: patients respond differently to the same treatment due to individual differences in genetics, physiology, and medical history. Traditional “one-size-fits-all” approaches fail to capture this heterogeneity, limiting treatment effectiveness.^{1–4} A promising solution involves developing individualized treatment regimes (ITRs), which assign treatments based on individual characteristics to optimize some functional of the counterfactual outcome distribution in the target population, commonly referred to as the value function. The objective is to learn a decision rule that maximizes this value function, thereby selecting the best treatment among available options for each patient. The optimal ITR is the rule that attains the maximal value, defined as the supremum of the value function over a specified class of decision rules. This methodology relies on causal identification assumptions (see Section 2 for details) and has transformed healthcare decision-making,^{5–9} with applications in economics^{10–13} and computer science.^{14,15}

¹Department of Statistics and Data Science, Tsinghua University, Beijing, China

²School of Statistics, Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China

³Beijing International Center for Mathematical Research, Peking University, Beijing, China

⁴Department of Statistics and Data Science, Duke-NUS Graduate Medical School, National University of Singapore, Singapore, Singapore

#These authors contributed equally to this work.

Corresponding author:

Yong Zhou, School of Statistics, Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China.
Email: yzhou@fem.ecnu.edu.cn

Statistical methods for estimating optimal treatment decisions are commonly categorized into two classes: model-based methods and direct search methods. Model-based approaches for estimating optimal ITRs encompass several well-established strategies that rely on modeling the outcome or treatment effect. Among them, Q-learning is a widely used method^{16–18} that estimates the so-called Q-function, which represents the expected outcome as a function of patient characteristics and treatment, and then derives the treatment rule that yields the highest expected outcome. Another example is A-learning, which focuses on modeling treatment contrasts directly and constructs optimal rules based on estimated treatment effects.^{19–21} On the other hand, direct search methods aim to directly maximize the value function to obtain the optimal ITR. Early work in this direction includes²² and²³ who applied inverse probability weighting (IPW) to estimate the value function and then leveraged classification techniques to learn the optimal ITR. To address potential model misspecification, augmented IPW (AIPW) methods have been developed to achieve double robustness.^{24–27} More recently, cross-fitting techniques borrowed from semiparametric statistics²⁸ have been incorporated into ITR learning,^{11,29} enabling the use of flexible machine learning methods for estimating nuisance functions without sacrificing the efficiency of the resulting ITR estimator. A comprehensive review of ITR learning methods can be found in Kosorok and Laber³ and references therein.

While these methods have demonstrated success in conventional outcome settings, estimating optimal ITRs with survival data presents substantial challenges. Early works in survival-based ITR estimation include Goldberg and Kosorok³⁰ and Zhao et al.,⁹ who extended Q-learning and doubly robust methods to accommodate censored survival outcomes. Subsequent developments have further advanced this research frontier, including Jiang et al.³¹ and Díaz et al.,³² followed by more recent innovations such as dynamic treatment regime estimation,³³ imputation-based Q-learning methods,³⁴ multi-stage regimes with dependent censoring,³⁵ and transfer learning approaches for censored survival data.³⁶ However, these existing approaches typically assume complete survival outcome information for all subjects, which is often unrealistic in practice where large amounts of covariate data may be available but survival outcome ascertainment is limited by cost and time constraints. While semi-supervised learning has shown promise in various statistical applications, only a limited number of studies have considered such data structures in optimal ITR estimation,^{37–39} and these are primarily designed for non-survival outcomes. The intersection of semi-supervised learning and survival-based ITR estimation remains largely unexplored, motivating the development of methods that can effectively leverage both labeled and unlabeled data in the presence of survival data.

To address the challenges outlined above, we propose a novel semi-supervised framework for estimating optimal ITRs in the presence of censored survival outcomes and partially labeled data. We begin by establishing a supervised baseline using inverse probability weighting to estimate the value function under censoring. Building on this, we introduce a semi-supervised estimation strategy that leverages robust imputation to safely incorporate unlabeled data, thereby enhancing statistical efficiency through variance reduction. To overcome the non-smoothness introduced by the indicator function in treatment rule estimation, we employ a kernel smoothing approach, which improves both numerical stability and the convergence rate of the ITR estimator. This smoothing yields substantial computational gains compared to traditional methods such as genetic algorithms, while also enabling theoretical guarantees. Specifically, we show that the proposed smoothed estimators are asymptotically normal, and that the convergence rate of the parameters indexing the optimal ITR can be improved from $n^{-1/3}$ to nearly $n^{-2/5}$. For valid inference, we adopt a resampling procedure, and demonstrate the effectiveness of our method through extensive simulation studies across diverse settings.

The remainder of this article is organized as follows. We first develop supervised and semi-supervised estimators for the optimal ITR with survival data and construct the corresponding kernel-smoothed estimators. We then establish the theoretical properties of the proposed estimators, demonstrating the superiority of our semi-supervised approach over the supervised method. The numerical studies section presents Monte Carlo simulation studies and an application to the METABRIC dataset. We conclude with a discussion of the strengths and limitations of the proposed methodology. Proofs of asymptotic properties and additional numerical results are provided in the Appendix.

2 Methodology

Let $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ be the predictor vector with bounded support $\mathcal{X}$ and fixed dimensionality p , and let A be the binary treatment taking values in $\mathcal{A} = \{0, 1\}$. Let $T \in \mathbb{R}^+$ denote the survival time and $C \in \mathbb{R}^+$ denote the random censoring time. Denote $Y = \min(T, C)$ as the observed survival time or the time to event, which is continuous, and $\Delta = I(T \leq C)$ as the censoring indicator.

We then use the potential outcomes framework,^{40,41} denoting $T^*(a)$ as the survival time had the subject were received treatment $a \in \mathcal{A}$. Following van der Laan and Robins,⁴² the causal estimand $E[T^*(a)]$ is nonparametrically identified under the following standard assumptions:

- A1. Consistency: If $A = a$, then $T^*(a) = T$ almost surely;
- A2. Positivity: $\Pr(A = a | \mathbf{X}) > 0$ almost surely for all $a \in \mathcal{A}$;
- A3. Strong ignorability : $A \perp \{T^*(1), T^*(0)\} | \mathbf{X}$;
- A4. Conditionally independent censoring: $C \perp \{T^*(1), T^*(0)\} | \mathbf{X}, A$.

Our primary objective is to identify and estimate an ITR $d(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{A}$, which is a mapping from the covariate space $\mathcal{X}$ to the treatment space $\mathcal{A}$, that maximizes the expected outcome in a counterfactual world had this ITR been implemented. Let $\mathcal{D}$ denote the class of candidate ITRs of interest. We define the potential outcome $T^*(d)$ under any $d \in \mathcal{D}$ as

$$T^*(d) = d(\mathbf{X})T^*(1) + \{1 - d(\mathbf{X})\}T^*(0),$$

and the value function⁴³ of d is defined by

$$M(d) = E[T^*(d)].$$

The optimal ITR is then defined as

$$d^{\text{opt}} = \arg \max_{d \in \mathcal{D}} M(d).$$

We will focus on the class of linear treatment regimes as in the literature,^{44–47} which offers a simple structure and good interpretability. This decision class is $\mathcal{D} = \{d_\beta(\mathbf{X}) = I(\beta' \mathbf{X} > 0) : \beta \in \mathcal{B}\}$, where $\mathcal{B} = \{\beta : \beta \in \mathbb{R}^p, \|\beta\| = 1\}$ and $\|\cdot\|$ is the Euclidean norm. Then we may determine d^{opt} by estimating the optimal coefficient

$$\beta_0 = \arg \max_{\beta \in \mathcal{B}} E[T^*(d_\beta)],$$

and obtain $d^{\text{opt}}(\mathbf{X}) = I(\beta_0' \mathbf{X} > 0)$.

2.1 Supervised estimation

In the supervised learning setting, we observe n independent and identically distributed (iid) samples

$$\mathcal{L} = \{(Y_i, \Delta_i, \mathbf{X}_i, A_i) : i = 1, 2, \dots, n\}.$$

Under the consistency assumption and the definition of Y , when $A = d_\beta(\mathbf{X})$ and $\Delta = 1$, the observed outcome is $Y = T^*(d_\beta(\mathbf{X}))$. This leads us to define an induced missing data structure^{26,48} with the missing data indicator $R(\beta)$ given by

$$R(\beta) = [Ad_\beta(\mathbf{X}) + (1 - A)\{1 - d_\beta(\mathbf{X})\}]\Delta.$$

In that sense, the potential outcome $T^*(d_\beta(\mathbf{X}))$ we interested can be observed only when $R(\beta) = 1$ and ‘missing’ otherwise. Then the observed data from labeled source could be represented as $\mathcal{L} = \{(\mathbf{X}_i, R_i(\beta)Y_i, R_i(\beta)) : i = 1, 2, \dots, n\}$. Furthermore, we denote the conditional survival function $P(C \geq t | \mathbf{X}, A)$ as $S_C(t | \mathbf{X}, A)$. The induced missing data structure satisfies the missing at random (MAR) assumption.

Let $\pi_R(\beta) = P(R(\beta) = 1 | \mathbf{X}, T^*(0), T^*(1))$ be the conditional probability of ‘missing’ and $\pi_A(\mathbf{X}) = P(A = 1 | \mathbf{X})$ be the propensity score for the treatment. Zhou et al.⁴⁸ showed that when $R(\beta) = 1$, we have

$$\pi_R(\beta) = [\pi_A(\mathbf{X})d_\beta(\mathbf{X}) + (1 - \pi_A(\mathbf{X}))\{1 - d_\beta(\mathbf{X})\}]S_C(Y | \mathbf{X}, A).$$

Based on the inverse probability weighting principle, the value function under a treatment rule d_β is given by

$$M(\beta) = E \left[\frac{R(\beta)}{\pi_R(\beta)} Y \right]. \quad (1)$$

When the propensity score $\pi_A(\mathbf{X})$ is unknown, it can be estimated consistently using either parametric models, such as logistic regression,⁴⁹ or nonparametric approaches, such as generalized random forests.⁵⁰ To isolate the main ideas from estimation complexities, we focus on the randomized trial setting where $\pi_A(\mathbf{X}) = 0.5$. In this case, the weighting function

simplifies to $\pi_R(\beta) = (1/2)S_C(Y | \mathbf{X}, A)$, which is independent of β . The conditional survival function $S_C(Y | \mathbf{X}, A)$ can be estimated using the Kaplan–Meier estimator,^{51,52} denoted by $\hat{S}_C(Y | \mathbf{X}, A)$. The resulting estimator of the value function is

$$\hat{M}^L(\beta) = P_n \left[\frac{2R(\beta)}{\hat{S}_C(Y | \mathbf{X}, A)} Y \right], \quad (2)$$

where $P_n f = (1/n) \sum_{i=1}^n f(X_i)$ denotes the empirical expectation. The corresponding estimator for the optimal parameter β_0 is obtained by

$$\hat{\beta}_L = \arg \max_{\beta \in B} \hat{M}^L(\beta).$$

To facilitate computation, we focus on the terms in (2) that depend on β , omitting constants irrelevant to the maximization. This leads to the equivalent optimization problem

$$\hat{\beta}_L = \arg \max_{\beta \in B} \hat{V}^L(\beta),$$

where the simplified objective function is defined as

$$\hat{V}^L(\beta) = P_n \left[\frac{(2A - 1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} d_\beta \right], \quad (3)$$

with $d_\beta = I(\beta^\top \mathbf{X} > 0)$ denoting a binary decision rule in the function class $\mathcal{D}$.

2.2 Semi-supervised estimation

In the semi-supervised (SS) setting, we have access to an additional unlabeled dataset

$$\mathcal{U} = \{(X_i, A_i) : i = n + 1, n + 2, \dots, n + N, N \geq 1\}$$

consisting of N iid observations without outcome information, which is independent of $\mathcal{L}$. Following the traditional SS framework,⁵³ we assume that observations in both $\mathcal{L}$ and $\mathcal{U}$ are drawn from the same underlying distribution and the n labeled observations are randomly selected from the combined dataset, which implies that the missingness of Y in the unlabeled observations is missing completely at random.(MCAR;^{54,55})

To fully leverage the auxiliary dataset $\mathcal{U}$, the following semi-supervised estimator is proposed. We note the following equivalent formulations:

$$\begin{aligned} & \arg \max_{\beta \in B} E \left[\frac{R(\beta)}{\pi_R(\beta)} Y \right] \\ &= \arg \max_{\beta \in B} E \left[\frac{(2A - 1)\Delta Y}{S_C(Y | \mathbf{X}, A)} d_\beta \right] \\ &= \arg \max_{\beta \in B} E [\mu(\beta' \mathbf{X}) d_\beta] + E \left[\left\{ \frac{(2A - 1)\Delta Y}{S_C(Y | \mathbf{X}, A)} - \mu(\beta' \mathbf{X}) \right\} d_\beta \right], \end{aligned} \quad (4)$$

where $\mu(\beta' \mathbf{X})$ represents the conditional expectation $E[\{((2A - 1)\Delta Y)/(S_C(Y | \mathbf{X}, A))\} | \beta' \mathbf{X}]$. Similar to formula 2.1 in Chakraborty et al.,⁵⁶ the second term in (4) ensures global robustness with respect to any working model posited to estimate the conditional expectation $\mu(\beta' \mathbf{X})$. Various estimation methods, such as parametric models, smoothing splines, and kernel smoothing, can be employed for this purpose. In this context, we focus on the kernel smoothing estimator for $\mu(\beta' \mathbf{X})$, which is given by

$$\hat{\mu}(\beta' \mathbf{X}) = \frac{\sum_{i=1}^n K_h(\beta' \mathbf{X}_i - \beta' \mathbf{X}) \hat{O}_i}{\sum_{i=1}^n K_h(\beta' \mathbf{X}_i - \beta' \mathbf{X})},$$

where $\hat{O}_i = ((2A_i - 1)\Delta_i Y_i)/(\hat{S}_C(Y_i | \mathbf{X}_i, A_i))$ and $K_h(u - v) = (1/h)K((u - v)/h)$ with K being a kernel function and h being the bandwidth. The semi-supervised estimator for β_0 utilizes both labeled and unlabeled data $\mathcal{L} \cup \mathcal{U}$ is constructed

as

$$\begin{aligned}\hat{\beta}_{SS} &= \arg \max_{\beta \in B} \hat{V}^{SS}(\beta) \\ &= \arg \max_{\beta \in B} P_{n+N} [\hat{\mu}(\beta' \mathbf{X}) d_\beta] + P_n \left[\left\{ \frac{(2A-1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} - \hat{\mu}(\beta' \mathbf{X}) \right\} d_\beta \right].\end{aligned}\quad (5)$$

Remark 1. It is worth noting that this framework does not require the model to be correctly specified or even to exactly match the conditional expectation $\mu(\beta' \mathbf{X})$. However, to improve the estimation efficiency, we still attempt to estimate the conditional expectation. Details will be discussed in Section 3.

Remark 2. Constructing the imputation function with the conditional expectation $E[(2A-1)\Delta Y / (S_C(Y | \mathbf{X}, A)) | \mathbf{X}]$, that conditions fully on the multivariate $\mathbf{X}$ rather than on the single index $\beta' \mathbf{X}$, could achieve the semi-parametric efficiency bound for estimating β_0 . However, in the practical implementation, the multidimensional kernel smoothing suffers from the curse of dimensionality. To address this, we choose to adopt the dimension reduction approach. And in terms to the efficiency, when $\mu(\beta' \mathbf{X}) = E[(2A-1)\Delta Y / (S_C(Y | \mathbf{X}, A)) | \mathbf{X}]$, such as the conditional expectation $E[(2A-1)\Delta Y / (S_C(Y | \mathbf{X}, A)) | \mathbf{X}]$ admits a single-index model, our semi-supervised estimator remains asymptotically optimal. Otherwise, our method still outperforms traditional supervised methods.

Remark 3. Due to the sharp-edge effect⁵⁷ caused by the indicator function $I(\beta' \mathbf{X} > 0)$, we consider replacing it with alternative smooth functions such as the sigmoid function,⁵⁸ the hinge loss function^{22,59} and the kernel smoothing function,⁴⁷ which provide smoothed approximations to the indicator function. We primarily use kernel smoothing technique in our implementation, which is discussed in Section 2.3.

Remark 4. The proposed SS estimator is closely related to, yet fundamentally different from, existing doubly robust (DR) methods for optimal ITR learning with censored data (e.g., Zhao et al.⁹). Both approaches augment the inverse probability weighting (IPW) term with a survival outcome regression component. However, their roles and objectives differ. In classical DR methods, the outcome model is introduced to protect against misspecification of the censoring model, whereas in our SS framework it serves as an imputation function whose primary purpose is to leverage abundant unlabeled data to stabilize estimation and reduce variance. Moreover, standard DR estimators rely exclusively on labeled data, while our method incorporates unlabeled data through the leading term $P_{n+N}[\hat{\mu}(\beta' \mathbf{X}) d_\beta]$. Although the SS approach does not improve the convergence rate of the linear decision rule parameters, it achieves substantial variance reduction (Theorem 2), leading to notable efficiency gains in finite samples, as demonstrated in Appendix C. This highlights a key advantage of our method over classical DR approaches.

2.3 Kernel-smoothed estimation

In this section, the smoothed estimations are proposed. Specifically, the decision indicator function $d_\beta = I(\beta' \mathbf{X} > 0)$ is replaced by the smoothed decision function $\tilde{d}_\beta = \Phi((\beta' \mathbf{X})/\tilde{h}_n)$ in the value function estimations (3) and (5). Here, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution $N(0, 1)$, and $\tilde{h}_n$ is a sequence of bandwidths satisfying $\tilde{h}_n = o(n^{-1/5})$. Following Silverman,⁶⁰ $\tilde{h}_n$ was set as $\tilde{h}_n = 0.9n^{-0.2} \min\{\text{std}(\beta' \mathbf{X}), \text{IQR}(\beta' \mathbf{X})/1.34\}$, where ‘std’ denotes the standard deviation function and ‘IQR’ represents the interquartile range. Consequently, the smoothed versions of the supervised and semi-supervised estimators for β_0 can be respectively constructed as follows:

$$\tilde{\beta}_L = \arg \max_{\beta \in B} \tilde{V}^L(\beta), \quad \tilde{\beta}_{SS} = \arg \max_{\beta \in B} \tilde{V}^{SS}(\beta),$$

where

$$\begin{aligned}\tilde{V}^L(\beta) &= P_n \left[\frac{(2A-1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} \tilde{d}_\beta \right], \\ \tilde{V}^{SS}(\beta) &= P_{n+N} [\hat{\mu}(\beta' \mathbf{X}) \tilde{d}_\beta] + P_n \left[\left\{ \frac{(2A-1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} - \hat{\mu}(\beta' \mathbf{X}) \right\} \tilde{d}_\beta \right].\end{aligned}$$

Remark 5. In the practical implementation, it is difficult to optimize $\tilde{V}^{SS}(\beta)$ directly since the parameter β is also present in the non-linear kernel estimator. Thus we consider plugging in the initial consistent estimator $\tilde{\beta}_{init}$ of β obtained from the labeled data $\mathcal{L}$ in the kernel estimator $\hat{\mu}(\beta'X)$, and estimate β by iterating

$$\tilde{\beta}_{new} = \arg \max_{\beta \in \mathcal{B}} \tilde{V}^{SS}(\beta, \hat{\mu}(\tilde{\beta}'_{old}X)),$$

where

$$\tilde{V}^{SS}(\beta, \hat{\mu}(\tilde{\beta}'_{old}X)) = P_{n+N} [\hat{\mu}(\tilde{\beta}'_{old}X) \tilde{d}_\beta] + P_n \left[\left\{ \frac{(2A-1)\Delta Y}{\hat{S}_C(Y|X, A)} - \hat{\mu}(\tilde{\beta}'_{old}X) \right\} \tilde{d}_\beta \right]$$

with a little abuse of notation.

A natural choice of the initial value of $\tilde{\beta}_{old}$ is the smoothed supervised estimator $\tilde{\beta}_L$. We finally denote the estimator $\tilde{\beta}_{new}$ at convergence as our smoothed semi-supervised estimator $\tilde{\beta}_{SS}$ of β .

For the above non-convex optimization problem, we use the proximal gradient descent algorithm proposed by Nesterov,⁶¹ which was also adopted in Wu and Wang.⁴⁷ Some notes on the convergence of this algorithms can be found in Appendix B. Let $\nabla_m f(\beta)$ denote the m -th partial derivative operator of $f(\beta)$ with respect to β . For notation simplicity, the subscript m is omitted when $m = 1$. A sequence of iterates $\beta^{(t)}$ (where ‘ $\cdot$ ’ represents the corresponding superscript (or subscript) ‘L’ or ‘SS’ indicating the type of estimation) is generated by

$$\beta^{(t)} = \arg \min_{\beta \in \mathcal{B}} \left\{ -\tilde{V}(\beta^{(t-1)}) - \langle \nabla \tilde{V}(\beta^{(t-1)}), \beta - \beta^{(t-1)} \rangle + r_t \|\beta - \beta^{(t-1)}\|^2 \right\},$$

where $\langle \cdot, \cdot \rangle$ represents the inner product and the choice of r_t could be found in Fan et al.⁶² Then the closed form solution can be written as

$$\beta^{(t)} = \beta^{(t-1)} + \frac{\nabla \tilde{V}(\beta^{(t-1)})}{2r_t}.$$

The updating program stops when there is no improvement of the value $V(\beta)$, that is

$$\tilde{V}(\beta^{(t)}) < \tilde{V}(\beta^{(t-1)}) + \langle \nabla \tilde{V}(\beta^{(t-1)}), \beta^{(t)} - \beta^{(t-1)} \rangle - r_t \|\beta^{(t)} - \beta^{(t-1)}\|^2.$$

The specific forms of the gradient for the supervised and semi-supervised estimations of the value function are given by

$$\begin{aligned} \nabla \tilde{V}^L(\beta^{(t)}) &= P_n \left[\frac{(2A-1)\Delta Y}{\hat{S}_C(Y|X, A)} \phi \left(\frac{\beta^{(t)'}X}{\tilde{h}_n} \right) \frac{X}{\tilde{h}_n} \right], \\ \nabla \tilde{V}^{SS}(\beta^{(t)}) &= P_{n+N} \left[\hat{\mu}(\beta^{(t-1)'}X) \phi \left(\frac{\beta^{(t)'}X}{\tilde{h}_n} \right) \frac{X}{\tilde{h}_n} \right] \\ &\quad + P_n \left[\left\{ \frac{(2A-1)\Delta Y}{\hat{S}_C(Y|X, A)} - \hat{\mu}(\beta^{(t-1)'}X) \right\} \phi \left(\frac{\beta^{(t)'}X}{\tilde{h}_n} \right) \frac{X}{\tilde{h}_n} \right] \end{aligned}$$

respectively, where $\phi(\cdot)$ denotes the probability density function of standard normal distribution $N(0, 1)$.

3 Statistical inference

3.1 Asymptotic properties

Let $V(\beta) = E[(2A-1)\Delta Y / (S_C(Y|X, A)) d_\beta]$ and β_0 be the maximizer of $V(\beta)$. To establish the asymptotic properties, the following regularity conditions are assumed.

- C1. For every $\varepsilon > 0$, $\sup_{\|\beta - \beta_0\| \geq \varepsilon} V(\beta) < V(\beta_0)$.
- C2. The density function of $\beta'X$ is everywhere positive for $\beta \in \mathcal{N}(\beta_0, \delta)$, where $\mathcal{N}(\beta_0, \delta)$ is a neighborhood of β_0 . And $\mathcal{X}$, the support of X , is not contained in a proper linear subspace of $\mathbb{R}^p$.

- C3. The censoring variable C has a continuously differentiable density function on $[0, L]$, where L is the end time of the study. And there exists a constant $c_1 > 0$ such that $S_C(\cdot) \geq c_1 > 0$, where $S_C(\cdot)$ denotes the survival function of C .
- C4. The kernel function $K(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is a symmetric probability density function. $K(\cdot)$ is twice continuously differentiable and Lipschitz and the second derivative satisfies the Lipschitz condition with

$$\int K^2(u)du < \infty, \mu_2(K) = \int u^2 K^2(u)du < \infty.$$

- C5. The bandwidth h in kernel function satisfies $h \rightarrow 0$, $nh \rightarrow \infty$ and $nh^4 \rightarrow 0$ as $n \rightarrow \infty$.

Conditions C1-C2 are assumed to ensure the consistency of estimators for β . Condition C1 implies that the value function $V(\beta)$ has a unique maximizer at β_0 . Condition C2 is assumed to demonstrate the continuity of value function. It holds true when there exists a component of $\mathbf{X}$, denoted as X_j , with an everywhere positive density conditional on the remaining covariates, and $\beta_{j0} \neq 0$. Condition C3 is a common assumption in survival analysis, where L represents the maximum follow-up time. If survival time exceeds L , it becomes unobservable, resulting in right-censoring of the observed survival times. Conditions C4-C5 are typically used for kernel estimation and are standard conditions for obtaining consistency of the semi-parametric estimator.⁶³

Theorem 1. Under conditions C1-C5, as $n \rightarrow \infty$ and $N \rightarrow \infty$, we have

- (a) $\tilde{\beta}_L \rightarrow \beta_0$ in probability.
- (b) $\tilde{\beta}_{SS} \rightarrow \beta_0$ in probability.

This theorem gives the consistency of the indexing parameter $\beta \in \mathcal{B}$ of ITR d_β . By Condition C1 and Theorem 5.7 in van der Vaart,⁶⁴ it is sufficient to show the uniform convergence of value function to get the consistency of the M-estimators. Proof details are given in the Appendix A.2.

Corollary 1. Under conditions C1-C5, as $n \rightarrow \infty$ and $N \rightarrow \infty$, we have

- (a) $\tilde{V}^L(\beta) \rightarrow V(\beta)$ in probability for any $\beta \in \mathcal{B}$.
- (b) $\tilde{V}^{SS}(\beta) \rightarrow V(\beta)$ in probability for any $\beta \in \mathcal{B}$.

Corollary 2. Under conditions C1-C5, as $n \rightarrow \infty$ and $N \rightarrow \infty$, we have

- (a) $\hat{\beta}_L \rightarrow \beta_0$ in probability.
- (b) $\hat{\beta}_{SS} \rightarrow \beta_0$ in probability.

Corollary 3. Under conditions C1-C5, as $n \rightarrow \infty$ and $N \rightarrow \infty$, we have

- (a) $\hat{V}^L(\beta) \rightarrow V(\beta)$ in probability for any $\beta \in \mathcal{B}$.
- (b) $\hat{V}^{SS}(\beta) \rightarrow V(\beta)$ in probability for any $\beta \in \mathcal{B}$.

Corollary 1, Corollary 2 and Corollary 3 illustrate the consistency of smoothed value functions, estimated parameters and corresponding value functions without smoothing respectively. These corollaries are immediate conclusions from the proof of Theorem 1.

To simplify the technical proof for asymptotic distributions of estimators, we adopt an equivalent identification condition for parameter that ‘there exists a covariate with a nonzero coefficient whose conditional distribution given the other covariates is continuous and its coefficient is normalized to have absolute value one’.⁴⁷ Throughout this section, we assume that β_1 , the coefficient of the first covariate X_1 , satisfies $|\beta_1| = 1$ without loss of generality. Let $\mathbf{X} = (X_1, \mathbf{X}_{-1}')'$ and $\beta = (\beta_{\cdot,1}, \beta_{\cdot,-1}')'$. The following conditions C6-C8 are assumed to derive the asymptotic normality of kernel-smoothed estimations.

- C6. Let $G(z, \mathbf{X}_{-1}) = E[T^*(1) - T^*(0)|z, \mathbf{X}_{-1}]$ with $z = \beta_0' \mathbf{X}$, $G^{(1)}(z, \mathbf{X}_{-1})$ and $G^{(2)}(z, \mathbf{X}_{-1})$ be the first and second partial derivative of $G(z, \mathbf{X}_{-1})$ respectively with respect to z , and $G(0, \mathbf{X}_{-1}) = 0$ for almost every $\mathbf{X}_{-1}$. For all z in a neighborhood of 0 and $i=1$ or 2, $G^{(i)}(z, \mathbf{X}_{-1})$ is a continuous function of z satisfying $|G^{(i)}(z, \mathbf{X}_{-1})| < C_G$ with a positive constant C_G for almost every $\mathbf{X}_{-1}$.

- C7. Let $f(z|\mathbf{X}_{-1})$ be the density function of z conditional on $\mathbf{X}_{-1}$. And for all z in a neighborhood of 0, $f^{(1)}(z|\mathbf{X}_{-1})$, the first partial derivative with respect to z , is a continuous function of z and uniformly bounded by a positive constant C_f for almost every $\mathbf{X}_{-1}$.
- C8. $V(\beta)$ is twice continuously differentiable on an open neighborhood of β_0 and the Hessian matrix $\nabla_2 V(\beta_0)$ is negative definite.

Following the consistency derived in Theorem 1 and the identification condition $|\beta_1| = 1$, we have $P(\tilde{\beta}_{L,1} = \beta_{0,1}) \rightarrow 1$ and $P(\tilde{\beta}_{SS,1} = \beta_{0,1}) \rightarrow 1$ as $n, N \rightarrow \infty$. Thus we mainly focus on the asymptotic normality for $\tilde{\beta}_{L,-1}$ and $\tilde{\beta}_{SS,-1}$.

Theorem 2. Let $a_1 = \int \zeta \dot{\phi}(\zeta) d\zeta$, $a_2 = \int \phi^2(\zeta) d\zeta$. Denote the following notations that $Q = a_1 E[\mathbf{X}_{-1} \mathbf{X}_{-1}' G^{(1)}(0, \mathbf{X}_{-1}) f(0|\mathbf{X}_{-1})]$, $D = (1/2) a_2 E[\mathbf{X}_{-1} \mathbf{X}_{-1}' E[T^*(1)^2 + T^*(0)^2 | z = 0, \mathbf{X}_{-1}] f(0|\mathbf{X}_{-1})]$ and $D^* = a_2 E[\mu^2(0) \mathbf{X}_{-1} \mathbf{X}_{-1}' f(0|\mathbf{X}_{-1})]$. Under conditions C1–C8 and $\tilde{h}_n = o(n^{-1/5})$, we have

- (a) $\sqrt{n\tilde{h}_n}(\tilde{\beta}_{L,-1} - \beta_{0,-1}) \rightarrow N(0, \Sigma_L)$ in distribution,
- (b) $\sqrt{n\tilde{h}_n}(\tilde{\beta}_{SS,-1} - \beta_{0,-1}) \rightarrow N(0, \Sigma_L - (1 - \rho)\Sigma_S)$ in distribution,

where $\Sigma_L = Q^{-1} D Q^{-1}$, $\Sigma_S = Q^{-1} D^* Q^{-1}$, $\rho \in [0, 1)$ is a constant such that $\rho_{n,N} = n/n + N \rightarrow \rho$ as $n, N \rightarrow \infty$.

This theorem establishes the asymptotic normality of the estimators for the index parameters with a convergence rate $O_p((n\tilde{h}_n)^{-1/2})$, which can be arbitrarily close to $n^{-2/5}$. Theorem 2 shows that the asymptotic variance of semi-supervised estimator is no greater than that of the supervised estimator for $\rho \in [0, 1)$. This improvement becomes more pronounced as ρ decreases. Specifically, when n is fixed and N increases, $\rho_{n,N}$ approaches 0, leading to a corresponding decrease in the asymptotic variance of $\tilde{\beta}^{SS}$. When both n and N go to infinity, $\rho_{n,N}$ can also approach 0 (a common assumption in the semi-supervised framework,⁵⁴) which violates the “positivity assumption” commonly required in the missing data literature⁶³ that ρ is bounded away from zero. However, it is precisely this violation that enables the variance reduction in our framework. We further demonstrate these theoretical findings through numerical simulations in Section 4.1. The proof of Theorem 2 is given in Appendix A.3.

To further illustrate the advantages of the smoothing technique in terms of convergence rate, we provide the following theorem.

Theorem 3. Under conditions C1–C5 and C8, we have

- (a) $|\hat{\beta}_L - \beta_0| = O_p(n^{-1/3})$,
- (b) $|\hat{\beta}_{SS} - \beta_0| = O_p(n^{-1/3})$.

Theorem 3 states that the convergence rate of the estimated parameters without smoothing is $O_p(n^{-1/3})$, which is slower than $O_p((n\tilde{h}_n)^{-1/2})$ under the assumption that $\tilde{h}_n = o(n^{-1/5})$. Detailed proofs are given in the Appendix A.4.

3.2 Resampling-based variance estimation

We utilize a straightforward resampling technique⁶⁵ by repeatedly perturbing the objective function to estimate the variance of supervised and semi-supervised estimators. This method has been widely used in the literature for inference.^{45,47,48,66} Let $\{\xi_i : i = 1, 2, \dots, n + N\}$ be a sequence of iid positive random weights independent of the data $\mathcal{L} \cup \mathcal{U}$, from the distribution with mean one and variance one such as the gamma distribution $Ga(1, 1)$. We consider the perturbed smoothed estimators that

$$\check{\beta}_L = \arg \max_{\beta \in B} \check{V}^L(\beta), \quad \check{\beta}_{SS} = \arg \max_{\beta \in B} \check{V}^{SS}(\beta),$$

where with $\check{\mu}(\beta' \mathbf{X}) = (\sum_{i=1}^n K_h(\beta' \mathbf{X}_i - \beta' \mathbf{X}) \hat{O}_i \xi_i) / (\sum_{i=1}^n K_h(\beta' \mathbf{X}_i - \beta' \mathbf{X}) \xi_i)$,

$$\check{V}^L(\beta) = P_n \left[\xi \frac{(2A-1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} \tilde{d}_\beta \right],$$

$$\check{V}^{SS}(\beta) = P_{n+N} [\xi \check{\mu}(\beta' \mathbf{X}) \tilde{d}_\beta] + P_n \left[\xi \left\{ \frac{(2A-1)\Delta Y}{\hat{S}_C(Y | \mathbf{X}, A)} - \check{\mu}(\beta' \mathbf{X}) \right\} \tilde{d}_\beta \right].$$

It is worth noting that the estimator $\hat{\mu}(\beta' \mathbf{X})$ does not need to be perturbed technically. However, in order to make more accurate standard error estimates with finite samples, we therefore perturb it by default.^{67,68}

Then we follow the proximal gradient descent algorithm as discussed in Section 2.3 and repeat the procedure for B times to get a sequence of estimators $\{\check{\beta}_\cdot^b : b = 1, \dots, B\}$. Similar to the arguments from Wu and Wang⁴⁷ and Jin et al.,⁶⁵ $\sqrt{n\check{h}_n}(\check{\beta}_{(\cdot)} - \beta_0)$ has the same asymptotic distribution as that of $\sqrt{n\check{h}_n}(\check{\beta}_{(\cdot)} - \check{\beta}_{(\cdot)})$. Thus we denote the empirical variance (or standard deviation) of $\{\check{\beta}_\cdot^b : b = 1, \dots, B\}$ as the estimator of population asymptotic variance (or standard deviation).

4 Numerical studies

4.1 Monte Carlo simulations

In this section, a series of simulation studies are conducted to evaluate the performance of our proposed smoothed semi-supervised estimation, comparing it to the corresponding smoothed supervised estimation. In case 1, we generate the covariates X_1 from the discrete distribution $P(X_1 = 1) = P(X_1 = -1) = 0.5$ and X_2 from $U(-1, 1)$. The response variable in the absence of censoring is generated by the accelerated failure time (AFT) model that $T = \exp\{\eta'_0 \mathbf{X} + A\beta'_0 \mathbf{X} + \epsilon\}$, with $\mathbf{X} = (1, X_1, \dots, X_p)$, treatment $A \sim \text{Bernoulli}(0.5)$ and error term $\epsilon \sim N(0, 1)$. We set $\eta_0 = (0, 0.25, -0.25)$ and $\beta_0 = (\beta_0, \beta_1, \beta_2) = (0, 1, 1)$ as the true parameters thus $p = 2$ here. Note that β_0 is the coefficient of the intercept, but what we are most interested in are the efficiency of (β_1, β_2) . The censoring time C follows an exponential distribution such that the observations have the corresponding censoring rate (cr). The observed response $Y = \min\{T, C\}$ with censoring indicator $\Delta = I(T \leq C)$. Thus, the true optimal treatment regime is given by $d^{opt}(\mathbf{X}) = I(\beta'_0 \mathbf{X} > 0)$. And we adopt the normalization $\beta_{std} = (\beta_0 / \|\beta_0\|)$ for identification. We fix the sample size $n=500$ for labeled data and vary the sample size of unlabeled data in $N = 1000, 2500$ or 5000 to show the impact of increasing N for fixed n under different censoring rate. We thus generate the labeled random sample set $\mathcal{L} = \{(Y_i, \Delta_i, X_i, A_i) : i = 1, 2, \dots, n\}$ and unlabeled random sample set $\mathcal{U} = \{(X_i, A_i) : i = n+1, n+2, \dots, n+N\}$. The Gaussian kernel is employed with bandwidth $h = 0.2n^{-1/3}$. We conduct 1000 simulation runs and the numerical results are reported in the following tables.

Table 1 reports the bias (Bias) of the estimator, the standard deviation (SD) of the estimator, the mean squared error (MSE) of the estimator, and the percentages of making correct decisions (PCDs) which is defined as

$$\text{PCD} = \frac{1}{1000} \sum_{i=1}^{1000} P(d_{\check{\beta}_\cdot^i}(\mathbf{X}) = d_{\beta_0}(\mathbf{X})) \times 100\%$$

with $P(d_{\check{\beta}_\cdot^i}(\mathbf{X}) = d_{\beta_0}(\mathbf{X})) = 1 - (1/n) \sum_{j=1}^n |d_{\check{\beta}_\cdot^i}(\mathbf{X}_j) - d_{\beta_0}(\mathbf{X}_j)|$ and $\check{\beta}_\cdot^i$ representing the supervised (Sup) estimator or the proposed semi-supervised (SS) estimator for β in the i -th simulation. The efficiency (EFF) of the SS estimator compared with the Sup estimator is measured by the relative ratio of the reduction in the MSE that

$$\text{EFF} = \frac{\text{MSE of Sup estimator} - \text{MSE of SS estimator}}{\text{MSE of Sup estimator}},$$

which is also reported in our simulation results.

The result for case 1 shows the obvious superiority of the semi-supervised approach over the supervised approach. It can be seen that SS estimator has nearly the same bias but significantly smaller SD compared to Sup estimator. Therefore, the MSE of SS estimator is smaller than that of Sup estimator and thus SS estimator has improved efficiency compared to Sup estimator. The proper use of unlabeled data makes the parameter estimation much more accurate, which is also reflected by the improvement of PCD. Furthermore, as shown in Table 1, there exists a direct correlation between the magnitude of the unlabeled sample size and the reduction in standard deviation of SS estimation. Consequently, while maintaining a

Table 1. The result for case 1 with different censoring rate.

cr		10%			20%			30%			40%			
Est	N	β_1	β_2	PCD	β_1	β_2	PCD	β_1	β_2	PCD	β_1	β_2	PCD	
Sup		Bias	0.044	−0.099	94.54%	0.022	−0.078	93.42%	−0.015	−0.056	91.18%	−0.029	−0.054	90.17%
		SD	0.145	0.153		0.157	0.150		0.192	0.162		0.213	0.180	
		MSE	0.0230	0.0332		0.0251	0.0286		0.0371	0.0294		0.0462	0.0353	
SS	1000	Bias	0.049	−0.095	95.27%	0.024	−0.070	94.19%	−0.020	−0.036	91.82%	−0.049	−0.012	90.24%
		SD	0.125	0.130		0.134	0.125		0.158	0.126		0.171	0.125	
		MSE	0.0180	0.0259		0.0185	0.0205		0.0254	0.0172		0.0316	0.0158	
	2500	EFF	21.5%	21.9%		26.3%	28.2%		31.6%	41.6%		31.5%	55.3%	
		Bias	0.042	−0.084	95.16%	0.025	−0.076	94.17%	−0.024	−0.029	91.87%	−0.049	−0.008	90.47%
		SD	0.122	0.124		0.130	0.119		0.152	0.119		0.165	0.119	
	5000	MSE	0.0166	0.0224		0.0173	0.0179		0.0237	0.0150		0.0296	0.0142	
		EFF	27.5%	32.5%		31.3%	37.4%		36.2%	48.9%		35.9%	59.7%	
		Bias	0.039	−0.079	95.23%	0.016	−0.055	94.19%	−0.026	−0.024	91.88%	−0.049	−0.005	90.65%
		SD	0.119	0.120		0.124	0.115		0.149	0.116		0.160	0.115	
		MSE	0.0157	0.0206		0.0156	0.0163		0.0229	0.0140		0.0280	0.0133	
		EFF	31.7%	37.8%		37.8%	43.2%		38.3%	52.2%		39.4%	62.5%	

fixed labeled sample size and the same censoring rate, this reduction results in a corresponding improvement in efficiency. Furthermore, the accuracy of the optimal ITR derived from SS estimation consistently exceeds the PCD of the optimal ITR derived from labeled data only. In addition, the increase of the censoring rate is accompanied by the decrease in the PCD of the optimal ITR. Nonetheless, it is worth noting that the SS estimation still has a significant advantage over the Sup estimation even when the censoring rate increases. It has an efficiency improvement of greater than 20% and maintains a higher PCD compared to Sup estimation. From the above discussion, it can be seen that although the supervised approach yields a good estimate of the optimal ITR, the semi-supervised approach can make this estimate more satisfactory by utilizing auxiliary information, which is very important in practical applications. This further demonstrates the validity of our approach.

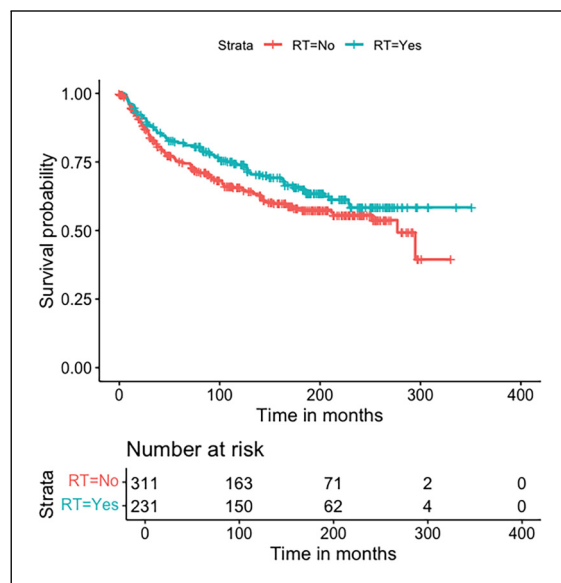
In the following, we show that our results are insensitive to the model for treatment-free effects. The true survival times without censoring are generated by $T = \exp\{\sin(\eta'_0 \mathbf{X}) + A\beta'_0 \mathbf{X} + \epsilon\}$ in case 2 and $T = \exp\{(\eta'_0 \mathbf{X})^2 + A\beta'_0 \mathbf{X} + \epsilon\}$ in case 3, respectively. The covariates $\mathbf{X}$, censoring time C and other parameters are set the same as that in case 1. The relevant simulation results are presented in Tables C.3 and Table C.4 in Appendix C. It is worth noting that the findings in case 2 and case 3 are consistent with those in case 1.

Because our imputation-based conditional expectation model utilizes single index projection for dimension reduction, we can consider higher-dimensional covariates when estimating optimal ITR, which is not feasible in Zhou et al.⁴⁸'s study. To investigate whether the semi-supervised method remains more efficient as the covariate dimension increases, we next consider case 4 and case 5 with $p = 4$ and $p = 6$, respectively, with $n = 500$, $N = 2000$, and $cr=10\%$ fixed for both cases. The covariate X_1 follows the Bernoulli distribution with probability 0.5, and $X_2, \dots, X_p$ independently follow $U(-1, 1)$. The true parameters in the AFT model built in case 1 for generating T are set as $\eta_0 = (0, 0.25, -0.25, 0.25, 0.25)$ and $\beta_0 = (0, 1, 0, 1, -1)$ in case 4 and $\eta_0 = (0, 0.25, -0.25, 0.25, 0.25, 0.25, 0.25, 0.25)$, $\beta_0 = (0, 0, 1, 1, 1, -1, 1)$ in case 5. Table 2 and C.5 present the corresponding simulation results of case 4 and case 5 respectively, including the Bias, SD, MSE, and EFF for the estimators of $\beta_1, \dots, \beta_p$, together with the corresponding PCD. The standard error (SE) estimated by $(1/1000) \sum_{i=1}^{1000} SE_i$ with every SE_i calculated based on 200 bootstrap samples as discussed in Section 3.2 and corresponding 95% empirical coverage probabilities (CP) for case 4 are also reported in Table 2. The SEs are close to SDs and observed empirical coverage probabilities are close to the nominal levels.

Comparing the simulation results in Table 1 with those in Table 2 and Table C.5, it is found that when the sample size and censoring rate are fixed, the computational complexity of parameter estimation increases significantly with the increase of the covariates dimension p . Additionally, the sample points become sparse in the high-dimensional space, which leads to a decrease in the accuracy of parameter estimation, and the efficiency of semi-supervised estimator also decreases. However, by utilizing the auxiliary information in unlabeled data, semi-supervised learning methods can mitigate the sparsity problem in multi-dimensional data to some extent. Therefore, despite the challenges, semi-supervised method remains superior to traditional supervised method, with the PCD being improved. In addition, to benchmark the proposed semi-supervised estimator against existing doubly robust methods for censored data,⁹ we conducted further simulation experiments, and the corresponding results are presented in Appendix C.

Table 2. The results for case 4.

Estimator		β_1	β_2	β_3	β_4	PCD
Sup	Bias	−0.031	−0.004	−0.038	0.055	87.89%
	SD	0.161	0.195	0.160	0.164	
	SE	0.157	0.188	0.158	0.156	
	MSE	0.0269	0.0380	0.0270	0.0299	
	CP	93.3%	93.4%	94.0%	93.3%	
SS	Bias	−0.023	0.002	−0.033	0.035	88.88%
	SD	0.132	0.167	0.137	0.135	
	SE	0.134	0.163	0.135	0.131	
	MSE	0.0180	0.0279	0.0199	0.0195	
	CP	95.3%	93.7%	92.2%	93.8%	
EFF		33.2%	26.7%	26.6%	35.0%	

**Figure 1.** Plot of the Kaplan–Meier estimator of survival functions of relapse free survival time in months for patients who received radio therapy (RT) or not, respectively.

4.2 METABRIC data example

To illustrate our kernel-smoothed semi-supervised method for optimal ITR estimating, we analyze the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset^{69–71} which is available at cbiportal website. This dataset contains clinical records of primary breast tumors from 2509 patients with long-term follow-up, including the patient's age at diagnosis, cellularity, estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status, neoplasm histologic grade, tumor size, Nottingham prognostic index, non-synonymous tumor mutational burden (TMB), chemotherapy, hormone therapy, type of breast surgery, radio therapy, overall and relapse free survival time (in months) with corresponding status, etc. In order to study the effect of *radio therapy* ($A = 1$, Yes; $A = 0$, No) on *relapse free survival time* Y in breast cancer patients, we considered the 542 patients who did not receive chemotherapy and hormone therapy, but only radio therapy, as our labeled data. Radio therapy is known to improve the relapse free survival time of patients.^{72–76} And intuitively, Figure 1 shows the estimated Kaplan–Meier survival curves of relapse free survival time in months for the groups of labeled patients who received and did not receive radio therapy, respectively.

However, we can not take a “one-size-fits-all” treatment regime for all patients because treatment decisions also depend on costs, side effects and other factors that affect an individual's quality of life.⁷⁷ Therefore we need to suggest appropriate treatment decisions for patients based on their individual characteristics.

Table 3. Estimated parameters indexing the optimal ITR for METABRIC.

Method	Sup			SS		
Predictors	Est	SE	p-value	Est	SE	p-value
intercept	−0.121	0.093	0.193	−0.110	0.063	0.079
age	−0.003	0.123	0.981	0.087	0.073	0.235
size	−0.096	0.100	0.341	−0.104	0.050	\textcolor{blue}{0.038}
TMB	0.082	0.157	0.603	0.312	0.134	\textcolor{blue}{0.020}
grade1	0.201	0.143	0.158	0.045	0.126	0.723
grade2	0.173	0.129	0.180	0.065	0.120	0.586
surgery	−0.948	0.074	\textcolor{blue}{0.000}	−0.931	0.067	\textcolor{blue}{0.000}

Numerical simulation results in Section 4.1 show that the increase in the dimension of covariates will lead to a decrease in the PCD of the optimal ITR, so the rational selection of covariates is very important. According to the conclusion of medical studies^{78,79} that receptors mainly respond to hormone therapy and HER2 targeted therapy, ER, PR and HER2 status are not included in the covariate set for studying the impact of radio therapy prognosis. Inspired by the extensive research in the medical literature on the effects of radio therapy on breast cancer patients,^{80–82} we select the following covariates **X** for our analysis: *age* at diagnosis, with a mean of 60 years old and a SD of 13 years old; tumor *size*, with a mean of 26 mm and an SD of 15 mm; *TMB* (non-synonymous), with a mean of 7 and an SD of 5; neoplasm histologic *grade*, coded as 1, 2, and 3, with 214 patients at grade 1, 976 patients at grade 2, and 1198 patients at grade 3; and the type of breast *surgery*, coded as -1 for breast-conserving surgery and 1 for mastectomy, with 785 patients undergoing breast-conserving surgery and 1170 patients undergoing mastectomy. For simplicity, the bolded parts of the covariates will be used to represent the corresponding covariates following. There were 542 patients who received only radio therapy as described above, and by removing individuals whose covariates were not fully observed and controlling for a balanced number of patients in the treatment and control groups, the final labeled dataset contains $n = 446$ observations with censoring rate about 40%. And the clinical information of $N = 1413$ patients whose covariates were fully observed in the remaining dataset was considered as the unlabeled dataset. To assess whether the MCAR assumption is tenable in the METABRIC data, we adopt the test approach proposed by Jamshidian and Jalal,⁸³ and implement it using the R package developed by Jamshidian et al.⁸⁴ Applying this nonparametric test to the METABRIC dataset yielded a p -value of 0.13. Thus, at the 0.05 significance level, there is not sufficient evidence to reject the null hypothesis that the outcome missingness mechanism satisfies MCAR. Before estimating procedure, we standardize the three continuous covariates “age,” “size,” and “TMB” by subtracting the mean and dividing by the standard deviation and use dummy variable encoding to represent the three-category variable ‘grade’ as two binary variables that grade1 (1, grade = 1; -1, grade $\neq$ 1) and grade2 (1, grade = 2; -1, grade $\neq$ 2). Table 3 reports the estimator (Est) of β which indexing the optimal ITR, the corresponding standard error (SE) estimated based on 200 bootstrap samples as discussed in Section 3.2, and p -value which is calculated by $2 - 2\Phi(|\text{Est}|/\text{SE})$. Significant p -value is marked blue.

Table 3 shows that both the supervised and semi-supervised methods identify the covariate surgery as significant at the 0.05 level, suggesting that breast cancer patients who undergo breast-conserving surgery should receive radio therapy. This result is in line with existing medical studies. For example, Sedlmayer et al.⁸⁵ showed that radio therapy after breast-conserving surgery improved survival for most patients,⁸⁶ showed that elderly patients who received radiation therapy after breast-conserving surgery had a lower risk of locoregional recurrence and⁸⁷ mentioned that radiotherapy after breast-conserving surgery for early breast cancer has become an indispensable part of breast cancer treatment, which can significantly reduce breast tumor recurrence and improve survival rate. Additionally, the semi-supervised method identifies two extra significant covariates ‘size’ and ‘TMB’. Thomas et al.⁸⁸ demonstrated that tumor size is a significant factor in predicting the recurrence risk in breast cancer patients after receiving radio therapy, with larger tumors indicating a higher risk of recurrence. In contrast, tumor grade is less critical than tumor size which does not show statistical significance in our analysis for this METABRIC dataset. ‘TMB’, which is often considered as an important biomarker in immunotherapy,⁸⁹ has an unclear relationship with radio therapy, despite existing research suggesting that local radio therapy can enhance anti-tumor immune responses.⁹⁰ Thus, our semi-supervised method offers a potentially viable direction for future research.

The semi-supervised method is more efficient than the supervised method, as also evidenced by the smaller estimation variance and a higher value function. The optimal individualized treatment recommendations derived from supervised and semi-supervised methods are presented in Table 4.

Under the optimal ITR obtained by the supervised method, the average recurrence-free survival time for breast cancer patients is 122.95 months. In contrast, the optimal ITR obtained by our semi-supervised method increases this average to

Table 4. Treatment recommendation for METABRIC study.

ITR	sup	SS
RT=Yes	752	762
RT=No	1107	1097

123.75 months. Figures C.1 and C.2 in the Appendix C show density plot comparisons of the five covariates grouped by whether or not radio therapy was recommended by the supervised and semisupervised methods, respectively. As can be seen in these plots, the covariate of breast-conserving surgery or mastectomy is significantly heterogeneous in both groups, resulting in the covariate ‘surgery’ playing a dominant role in determining our optimal ITR. The Sup and SS estimators generally recommend similar treatments for most patients. Only ten patients who are recommended not to receive radio therapy under the supervised method are re-recommended to receive radiotherapy under the semi-supervised method. For instance, among these ten patients, four are classified as grade 3. Although the four patients are all over 45 years old, their tumor sizes are not less than 20 mm. According to the DEGRO Practical Guidelines for Radiotherapy of Breast Cancer IV,⁹¹ patients with at least two risk factors (grade 3 and larger tumor size here) are strongly recommended to undergo radio therapy after mastectomy to reduce the long-term risk of recurrence. All these findings support that the proposed semi-supervised method is more efficient than the traditional supervised method in this real data application, which is attributed to the effective use of auxiliary unlabeled data.

5 Discussions

To conclude, we propose a novel semi-supervised learning method for estimating optimal ITR in right-censored survival data framework. Our research demonstrates that while both the semi-supervised (SS) estimator and the traditional IPW-based supervised estimator are unbiased, the SS estimator may have relatively smaller variance. Consequently, the proposed SS method significantly improves efficiency in the sense of MSE. We establish that the estimated parameters indexing the optimal ITR follow an asymptotic normal distribution. Extensive simulation studies in Section 4 indicate that, with a fixed number of labeled samples n , an increase in censoring rate or covariate dimension p could lead to a reduction in decision accuracy measured by PCD. Our proposed SS estimator consistently outperforms the supervised estimator. This advantage becomes more pronounced as the number of unlabeled samples N increases, aligning with the theoretical results discussed in Section 3. Furthermore, the kernel smoothing technique we applied to the indicator function in decision class $\mathcal{D}$ addresses the issue of sharp-edge, and improves the convergence rate of the parameters we interested in from $n^{-1/3}$ to arbitrarily close to $n^{-2/5}$. This kernel surrogate allows us to utilize the computationally efficient proximal gradient descent algorithm for our optimization problem. Additionally, the single-index projection dimension reduction method used in leveraging unlabeled data effectively mitigates the curse of dimension in nonparametric estimation, which is another feature of our proposed SS method.

Our methodology is developed under the assumption that the unlabeled outcomes are MCAR, which is standard in the semi-supervised learning literature^{54,55,92,93} and often adopted when labeling is primarily determined by study design. Nevertheless, in clinical applications, outcome availability may depend on observed covariates due to time or cost constraints, making the MCAR assumption potentially restrictive. To address this concern, our framework can be naturally extended to the more general missing at random (MAR) setting.⁹⁴ Let $S \in \{0, 1\}$ denote the source indicator, with $S = 1$ for labeled observations and $S = 0$ for unlabeled observations, and define $\pi_S(\mathbf{X}) = P(S = 1 \mid \mathbf{X})$. Under MAR, the semi-supervised value function can be modified as

$$\check{V}^{SS}(\beta) = P_{n+N} \left[\hat{\mu}(\beta' \mathbf{X}) d_\beta + \frac{S}{\pi_S(\mathbf{X})} \left\{ \frac{(2A - 1)\Delta Y}{\hat{S}_C(Y \mid \mathbf{X}, A)} - \hat{\mu}(\beta' \mathbf{X}) \right\} d_\beta \right].$$

When the missingness mechanism is MCAR, $\pi_S(\mathbf{X})$ reduces to the constant $n/(n + N)$, and the above expression coincides with the formulation used in this paper. This shows that our method can be viewed as a special case of a broader MAR framework, although a full theoretical investigation of this extension is beyond the scope of the current paper and will be pursued in future work.

Several extensions are also worth further investigation. First, our proposed method can be extended to estimate optimal ITRs under quantile criteria. While some existing studies have considered quantile-based criteria,^{48,95} none has explored such objectives under a semi-supervised survival data structure. Second, extending the current framework to multi-stage treatment problems for learning optimal dynamic treatment regimes (DTRs)^{21,96} under the semi-supervised setting is

another promising direction. Finally, when treatments are multi-class or continuous rather than binary,^{97,98} semi-supervised learning of optimal ITRs presents an intriguing avenue for future research.

ORCID iDs

Xintong Li  <https://orcid.org/0009-0009-7978-9836>

Jialiang Li  <https://orcid.org/0000-0002-9704-4135>

Yong Zhou  <https://orcid.org/0000-0001-9751-3141>

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Y. Zhou's work is supported by Natural Science Foundation of Shanghai (No. 23JS1400500), Shanghai Municipal Education Commission (No. 2024AI01002). J. Li's work was supported by Academic Research Funds A-8003585-00-0 and A-8001947-00-00 in Singapore.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

References

1. Imai K and Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann Appl Stat* 2013; **7**: 443–470.
2. Kent D, Steyerberg E and van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *Br Med J* 2018; **363**: k4245.
3. Kosorok MR and Laber EB. Precision medicine. *Annu Rev Stat Appl* 2019; **6**: 263–286.
4. Lipkovich I, Dmitrienko A and D'Agostino B. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med* 2017; **36**: 136–196.
5. Fu S, He Q, Zhang S, et al. Robust outcome weighted learning for optimal individualized treatment rules. *J Biopharm Stat* 2019; **29**: 606–624.
6. Jiang B, Song R, Li J, et al. Entropy learning for dynamic treatment regimes (with discussion). *Stat Sin* 2019; **29**: 1633–1710.
7. Li J, Li Y, Jin B, et al. Multi-threshold change plane model: estimation theory and applications in subgroup identification. *Stat Med* 2021; **40**: 3440–3459.
8. Liu P, Li J and Kosorok MR. Change plane model averaging for subgroup identification. *Stat Methods Med Res* 2023; **32**: 773–788.
9. Zhao YQ, Zeng D, Laber EB, et al. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* 2015; **102**: 151–168.
10. Athey S and Imbens GW. The state of applied econometrics: causality and policy evaluation. *J Econ Perspect* 2017; **31**: 3–32.
11. Athey S and Wager S. Policy learning with observational data. *Econometrica* 2021; **89**: 133–161.
12. Ding Y, Li Y and Song R. Statistical learning for individualized asset allocation. *J Am Stat Assoc* 2024; **119**: 639–649.
13. Imai K and van Dyk DA. Causal inference with general treatment regimes: generalizing the propensity score. *J Am Stat Assoc* 2004; **99**: 854–867.
14. Bennett A and Kallus N. Efficient policy learning from surrogate-loss classification reductions. In: *International conference on machine learning*, 2020, pp. 788–798. PMLR.
15. Shi C, Zhu J, Shen Y, et al. Off-policy confidence interval estimation with confounded Markov decision process. *J Am Stat Assoc* 2024; **119**: 273–284.
16. Qian M and Murphy SA. Performance guarantees for individualized treatment rules. *Ann Stat* 2011; **39**: 1180–1210.
17. Watkins C. *Learning from delayed rewards*. PhD Thesis, Cambridge University, Cambridge, England, 1989.
18. Watkins C and Dayan P. Q-learning. *Mach Learn* 1992; **8**: 279–292.
19. Murphy SA. Optimal dynamic treatment regimes. *J R Stat Soc Ser B: Stat Methodol* 2003; **65**: 331–355.
20. Robins JM, Hernan MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
21. Shi C, Fan A, Song R, et al. High-dimensional a-learning for optimal dynamic treatment regimes. *Ann Stat* 2018; **46**: 925–957.
22. Zhao Y, Zeng D, Rush AJ, et al. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc* 2012; **107**: 1106–1118.

23. Kitagawa T and Tetenov A. Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* 2018; **86**: 591–616.
24. Dudík M, Langford J and Li L. *Doubly robust policy evaluation and learning*. arXiv preprint arXiv:1103.4601, 2011.
25. Liu Y, Wang Y, Kosorok MR, et al. Augmented outcome-weighted learning for estimating optimal dynamic treatment regimes. *Stat Med* 2018; **37**: 3776–3788.
26. Zhang B, Tsiatis AA, Laber EB, et al. A robust method for estimating optimal treatment regimes. *Biometrics* 2012; **68**: 1010–1018.
27. Zhou X, Mayer-Hamblett N, Khan U, et al. Residual weighted learning for estimating individualized treatment rules. *J Am Stat Assoc* 2017; **112**: 169–187.
28. Bickel PJ. On adaptive estimation. *Ann Stat* 1982; **10**: 647–671.
29. Zhao YQ, Laber E, Ning Y, et al. Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *J Mach Learn Res* 2019; **20**: 1–23.
30. Goldberg Y and Kosorok MR. Q-learning with censored data. *Ann Stat* 2012; **40**: 529.
31. Jiang R, Lu W, Song R, et al. On estimation of optimal treatment regimes for maximizing t-year survival probability. *J R Stat Soc Ser B: Stat Methodol* 2017; **79**: 1165–1185.
32. Díaz I, Savenkov O and Ballman K. Targeted learning ensembles for optimal individualized treatment rules with time-to-event outcomes. *Biometrika* 2018; **105**: 723–738.
33. Simoneau G, Moodie EE, Nijjar JS, et al. Estimating optimal dynamic treatment regimes with survival outcomes. *J Am Stat Assoc* 2020; **115**: 1531–1539.
34. Lyu L, Cheng Y and Wahed AS. Imputation-based q-learning for optimizing dynamic treatment regimes with right-censored survival outcome. *Biometrics* 2023; **79**: 3676–3689.
35. Cho H, Holloway ST, Couper DJ, et al. Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring. *Biometrika* 2023; **110**: 395–410.
36. Zhao P, Josse J and Yang S. Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data. *J Mach Learn Res* 2025; **26**: 1–54.
37. Gunn K, Lu W and Song R. Adaptive semi-supervised inference for optimal treatment decisions with electronic medical record data. In: *Statistics in precision health: theory, methods and applications*, 2024, pp. 229–246. Springer.
38. Li X, Zhang S and Zhou Y. Semisupervised d-learning for optimal individualized treatment regimes. *Stat* 2025b; **14**: e70063.
39. Sonabend-W A, Laha N, Ananthakrishnan AN, et al. Semi-supervised off-policy reinforcement learning and value estimation for dynamic treatment regimes. *J Mach Learn Res* 2023; **24**: 1–86.
40. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; **66**: 688–701.
41. Splawa-Neyman J, Dabrowska DM and Speed TP. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Stat Sci* 1990; **5**: 465–472.
42. van der Laan MJ and Robins JM. *Unified methods for censored longitudinal data and causality*. Springer, 2003.
43. Manski CF. Statistical treatment rules for heterogeneous populations. *Econometrica* 2004; **72**: 1221–1246.
44. Chu J, Lu W and Yang S. Targeted optimal treatment regime learning using summary statistics. *Biometrika* 2023; **110**: 913–931.
45. Fan C, Lu W, Song R, et al. Concordance-assisted learning for estimating optimal individualized treatment regimes. *J R Stat Soc Ser B: Stat Methodol* 2017; **79**: 1565–1582.
46. Li C, Zeng D and Zhu W. A robust covariate-balancing method for learning optimal individualized treatment regimes. *Biometrika* 2025a; **112**: asae036.
47. Wu Y and Wang L. Resampling-based confidence intervals for model-free robust inference on optimal treatment regimes. *Biometrics* 2021; **77**: 465–476.
48. Zhou Y, Wang L, Song R, et al. Transformation-invariant learning of optimal individualized decision rules with time-to-event outcomes. *J Am Stat Assoc* 2023; **118**: 2632–2644.
49. Mo W and Liu Y. Efficient learning of optimal individualized treatment rules for heteroscedastic or misspecified treatment-free effect models. *J R Stat Soc Ser B: Stat Methodol* 2022; **84**: 440–472.
50. Athey S, Tibshirani J and Wager S. Generalized random forests. *Ann Stat* 2019; **47**: 1148–1178.
51. Gonzalez-Manteiga W and Cadarso-Suarez C. Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. *Commun Stat-Theory Methods* 1994; **4**: 65–78.
52. Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; **53**: 457–481.
53. Chapelle O, Schölkopf B and Zien A. *Semi-supervised learning*. MIT Press, 2006.
54. Chakraborty A and Cai T. Efficient and adaptive linear regression in semi-supervised settings. *Ann Stat* 2018; **46**: 1541–1572.
55. Song S, Lin Y and Zhou Y. A general m-estimation theory in semi-supervised framework. *J Am Stat Assoc* 2024; **119**: 1065–1075.
56. Chakraborty A, Dai G and Carroll RJ. *Semi-supervised quantile estimation: robust and efficient inference in high dimensional settings*. arXiv preprint arXiv:2201.10208, 2022.
57. Kim J and Pollard D. Cube root asymptotics. *Ann Stat* 1990; **18**: 191–219.

58. Fang EX, Wang Z and Wang L. Fairness-oriented learning for optimal individualized treatment rules. *J Am Stat Assoc* 2023; **118**: 1733–1746.
59. Liang S, Lu W, Song R, et al. Sparse concordance-assisted learning for optimal treatment decision. *J Mach Learn Res* 2018; **18**: 1–26.
60. Silverman BW. *Density estimation for statistics and data analysis*. London: Chapman & Hall, 1986.
61. Nesterov Y. Gradient methods for minimizing composite functions. *Math Program* 2013; **140**: 125–161.
62. Fan J, Liu H, Sun Q, et al. I-lamm for sparse learning: simultaneous control of algorithmic complexity and statistical error. *Ann Stat* 2018; **46**: 814.
63. Tsiatis AA. *Semiparametric theory and missing data*. Springer Series in Statistics. Springer New York, 2007.
64. van der Vaart AW. *Asymptotic statistics*. Vol. 3. Cambridge University Press, 2000.
65. Jin Z, Ying Z and Wei LJ. A simple resampling method by perturbing the minimand. *Biometrika* 2001; **88**: 381–390.
66. Peng L and Huang Y. Survival analysis with quantile regression models. *J Am Stat Assoc* 2008; **103**: 637–649.
67. Cheng D, Ananthakrishnan AN and Cai T. Robust and efficient semi-supervised estimation of average treatment effects with application to electronic health records data. *Biometrics* 2021; **77**: 413–423.
68. Cheng D, Chakraborty A, Ananthakrishnan AN, et al. Estimating average treatment effects with a double-index propensity score. *Biometrics* 2020; **76**: 767–777.
69. Curtis C, Shah SP, Chin SF et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; **486**: 346–352.
70. Pereira B, Chin SF, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun* 2016; **7**: 1–16.
71. Rueda OM, Sammut SJ, Seoane JA, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* 2019; **567**: 399–404.
72. Group EBCTC et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *The Lancet* 2005; **366**: 2087–2106.
73. Poleszczuk J, Luddy K, Chen L, et al. Neoadjuvant radiotherapy of early-stage breast cancer and long-term disease-free survival. *Breast Cancer Res* 2017; **19**: 1–7.
74. Rutqvist LE, Rose C and Cavallin-Ståhl E. A systematic overview of radiation therapy effects in breast cancer. *Acta Oncol (Madr)* 2003; **42**: 532–545.
75. Taylor C, Dodwell D, McGale P et al. Radiotherapy to regional nodes in early breast cancer: an individual patient data meta-analysis of 14 324 women in 16 trials. *The Lancet* 2023; **402**: 1991–2003.
76. Wöckel A, Wolters R, Wiegel T, et al. The impact of adjuvant radiotherapy on the survival of primary breast cancer patients: a retrospective multicenter cohort study of 8935 subjects. *Ann Oncol* 2014; **25**: 628–632.
77. Group EBCTC et al. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *The Lancet* 2000; **355**: 1757–1770.
78. Pegram M, Jackisch C and Johnston SR. Estrogen/her2 receptor crosstalk in breast cancer: combination therapies to improve outcomes for patients with hormone receptor-positive/her2-positive breast cancer. *NPJ Breast Cancer* 2023; **9**: 45.
79. Tommasi C, Airò G, Praticò F et al. Hormone receptor-positive/her2-positive breast cancer: Hormone therapy and anti-her2 treatment: an update on treatment strategies. *J Clin Med* 2024; **13**: 1873.
80. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med* 2002; **347**: 1233–1241.
81. Foulkes WD, Reis-Filho JS and Narod SA. Tumor size and survival in breast cancer—a reappraisal. *Nat Rev Clin Oncol* 2010; **7**: 348–353.
82. Fredholm H, Eaker S, Frisell J, et al. Breast cancer in young women: poor survival despite intensive treatment. *PLoS ONE* 2009; **4**: e7695.
83. Jamshidian M and Jalal S. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika* 2010; **75**: 649–674.
84. Jamshidian M, Jalal S and Jansen C. MISSMECH: an R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *J Stat Softw* 2014; **56**: 1–31.
85. Sedlmayer F, Sautter-Bühl ML, Budach W et al. Degro practical guidelines: radiotherapy of breast cancer i: radiotherapy following breast conserving therapy for invasive breast cancer. *Strahlenther Onkol* 2013; **189**: 825.
86. van de Water W, Bastiaannet E, Scholten AN, et al. Breast-conserving surgery with or without radiotherapy in older breast patients with early stage breast cancer: a systematic review and meta-analysis. *Ann Surg Oncol* 2014; **21**: 786–794.
87. Speers C and Pierce LJ. Postoperative radiotherapy after breast-conserving surgery for early-stage breast cancer: a review. *JAMA Oncol* 2016; **2**: 1075–1082.

88. Thomas F, Arriagada R, Mouriessse H, et al. Radical radiotherapy alone in non-operable breast cancer: the major impact of tumor size and histological grade on prognosis. *Radiother Oncol* 1988; **13**: 267–276.
89. O'Meara TA and Tolaney SM. Tumor mutational burden as a predictor of immunotherapy response in breast cancer. *Oncotarget* 2021; **12**: 394.
90. Jungles KM, Holcomb EA, Pearson AN, et al. Updates in combined approaches of radiotherapy and immune checkpoint inhibitors for the treatment of breast cancer. *Front Oncol* 2022; **12**: 1022542.
91. Wenz F, Sperk E, Budach W et al. Degro practical guidelines for radiotherapy of breast cancer iv. Radiotherapy following mastectomy for invasive breast cancer. *Strahlenther Onkol* 2014; **190**: 705–714.
92. Liu X, Wang T, Lin Y, et al. Semi-supervised inference for the high-dimensional quantile regression. *Sci China Math* 2026; **69**: 1033–1062.
93. Zhang Y and Bradic J. High-dimensional semi-supervised learning: in search of optimal inference of the mean. *Biometrika* 2022; **109**: 387–403.
94. Zhang Y, Chakraborty A and Bradic J. Double robust semi-supervised inference for the mean: selection bias under mar labeling with decaying overlap. *Inform Inferf: A J IMA* 2023; **12**: 2066–2159.
95. Wang L, Zhou Y, Song R, et al. Quantile-optimal treatment regimes. *J Am Stat Assoc* 2018; **113**: 1243–1254.
96. Zhang B, Tsiatis AA, Laber EB, et al. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* 2013; **100**: 681–694.
97. Ai C, Linton O, Motegi K, et al. A unified framework for efficient estimation of general treatment models. *Quant Econom* 2021; **12**: 779–816.
98. Kallus N and Zhou A. Policy evaluation and optimization with continuous treatments. In: *International conference on artificial intelligence and statistics*, 2018, pp.1243–1251. PMLR.